UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MASSACHUSETTS

DIRECT REPORT CORPORATION
d/b/a/ SHAREHOLDER.COM,

Plaintiff,

v.

CCBN.COM, INC., THE THOMSON CORPORATION, JOHN DOES 1 THROUGH 5, and JANE DOES 1 THROUGH 5,

Defendants.

Civil Action No. 04-10535-PBS

AFFIDAVIT OF JAMES BERRIMAN

1. I am employed by Goodwin Procter as Senior Counsel and Executive Manager of Litigation Technology. I am responsible for supervising electronic discovery projects and litigation-support operations. My professional experience in the field of automated litigation support began in 1982. I have been employed by Goodwin Procter since 1990 and I have held my present position since 1999.

2. As head of the firm's Litigation Technology Group, my duties include supervising an in-house team of ten e-discovery and litigation-support specialists. My duties also include selecting, supervising, and contracting with outside vendors for e-discovery and litigation-support services. As an attorney, I am also involved in making and responding to discovery requests that involve e-discovery. I have experience in negotiating the appropriate scope of e-discovery productions with adverse parties that have included major corporations, the DOJ, and the SEC.

3. I am familiar with the prevailing costs for e-discovery services. I am familiar with the prevailing industry-standard procedures for cost-effective retrieval of data from large collections of electronic documents, including collections stored on computer backup tapes.

4. One of the defining characteristics of e-discovery is that it typically involves enormous repositories of data -- far larger than traditional paper repositories -- that contain a low percentage of unique relevant documents. This is especially true for e-discovery involving backup tapes. The ratio of wheat to chaff tends to be very low while the overall quantity of data tends to be very large. There are several reasons for this, including: (a) the centralization of IT operations, which results in the aggregation of data for many employees on the same tapes regardless of relevance to the matter in dispute; (b) the archiving of system files and other non-user files, which are voluminous but typically irrelevant; (c) massive redundancy within each backup arising from the routine electronic distribution of files to many recipients within an organization; and (d) massive redundancy across multiple backups from the repeated periodic archiving of the same files.

5. When these factors are multiplied against each other, the volume of data in a collection of backup tapes can become almost inconceivably large, dwarfing anything that would exist in the paper world. For example, a single backup tape may typically contain 35 gigabytes of data. If that amount of data were printed out as plain text at 5,000 characters per page, the printed output would fill seven million pages. If the printed pages were packed into bankers boxes at 3,000 pages per box, the output of that single tape would fill more than 2,000 bankers boxes.

6. It is my understanding that in the present case, Plaintiff has requested the production of all communications between Defendant CCBN and 110 of CCBN's customers in

2002. In connection with that request, Plaintiff is now seeking to have Defendants restore the contents of all of the CCBN backup tapes in the possession of Defendants.

7. In connection with Plaintiff's request, I worked with the Defendants' IT Department to assess the scope and condition of the backup tape collection. The results of that assessment include the following:

(a) there are 478 backup tapes from CCBN that may contain, among many other things, email data from 1999 through 2004;

(b) on average, each backup tape is estimated to contain 35 gigabytes of uncompressed data, which means that the total amount of uncompressed data on all 478 tapes is estimated to be over 16 terabytes (or 16 thousand gigabytes);

(c) the backup system from which the tapes were generated and which was used to keep track of the contents of the tapes is no longer operational;

(d) there is no hardcopy index of the contents of the tapes;

(e) the tapes are not physically labeled sufficiently to determine the contents of particular tapes;

(f) the tapes are believed to contain consolidated weekly backups from various computer systems and servers, including the CCBN mail server as well as other systems not relevant to the discovery request;

(g) the backups are multi-tape backups, with each backup comprising a set of approximately five tapes containing approximately 175 gigabytes of uncompressed data, configured so that all tapes for each set must be present to restore a backup successfully;

(h) the size of the CCBN mail server during the relevant period is estimated to have been approximately 30 gigabytes;

(i) in light of (g) and (h), the percentage of data attributable to mail server backups would be approximately 15-20% of the total (i.e., approximately 30 gigabytes out of each 175 gigabyte backup set, or approximately 2-3 terabytes in total) and the remaining 80-85% would comprise files and systems other than the mail server;

(j) the 2-3 terabytes attributable to the mail server is believed to contain massive duplication due to the redundancy that occurs with weekly backups and the duplication that exists within each backup;

(k) the 2-3 terabytes attributable to the mail server is believed to contain many email accounts of people not relevant to this case and many emails not relevant to this case; and

(l) the 2-3 terabytes attributable to the mail server is believed to contain many emails outside the time period relevant to Plaintiff's request.

8. Based on this assessment, I believe that the quantity of unique, non-duplicative, responsive email data -- if any -- would be a tiny percentage of the entire 16-terabyte universe of data.

9. To determine the cost of complying with Plaintiff's request, I requested that an outside electronic discovery vendor, Electronic Evidence Discovery, Inc. ("EED"), provide an estimate. I have worked with EED on a number of past projects and I am familiar with their procedures and pricing. It is my understanding that EED is the oldest and most experienced electronic discovery vendor, with more than 18 years of experience. Due to the volume of work done by EED for clients of Goodwin Procter, EED provides our clients with reduced pricing.

10. EED prepared a "Proposal for Electronic Discovery Project", dated September 9, 2005, attached as Exhibit A. This contains the rates charged by EED for the various processes associated with e-discovery from backup tapes.

11. The first element of the EED estimate is for media restoration. The cost is $450 per tape, which includes $250 for cataloging (i.e., reading index data from a tape to determine its contents) and $200 for the restoration itself. The cataloging process is used to determine whether particular tapes are likely to contain data responsive to the request. Based on the catalog data, the typical approach is to select an appropriate subset of tapes that are likely to be responsive (e.g., containing backups for the relevant time period and system, such as the mail server) and not highly duplicative (e.g., quarterly or monthly backups rather than daily or weekly, since closely-spaced backups are highly duplicative and the small incremental increase in nonduplicative documents is greatly outweighed by the cost).

12. The next element of the EED estimate is for processing the restored data for online review. The cost is $1950 per gigabyte, but is applied only to the portions of the restored tapes that are actually extracted. Accordingly, the typical approach is to extract only the portions that are likely to contain data responsive to the request. For example, in the case of a backup tape for a mail server, this would involve extracting only the mailboxes of the relevant employees for the relevant period. It would not involve extracting and processing the entire contents of the mail server, since many employees and many emails are likely to be irrelevant to the case. Once these portions are extracted, duplicates are removed and the files are filtered by keywords to create a collection that contains the files that are most likely to be responsive to the request.

13. The next element of the EED estimate is for review management. The cost is $250 per hour for the time spent by the EED project manager in assisting the attorneys with formulating searches and controlling workflow.

14. The final element of the EED estimate is for exporting the production set following the review process. This includes exporting the data in a standard electronic format so the receiving party can load it into their litigation-support system (such as Summation or Concordance). The production set typically includes electronic images of the documents, email header data, and searchable text. The cost is 10 cents per "page" (i.e., per TIFF image equivalent to a printed page). The cost is applied only to the portion of the collection that is actually produced (i.e., the cost is not applied to the irrelevant documents). In the alternative, the production set can be printed to paper if the receiving party does not have a litigation-support system.

15. These processes reflect the prevailing industry standards for cost-effective production from backup tapes. In particular, they reflect multiple levels of culling designed to avoid the cost of restoring, extracting, reviewing, and producing irrelevant documents. The culling is designed to address the very low ratio of relevant documents typically found in a large data repository and to make e-discovery economically feasible.

16. In the present case, the collection of 478 tapes includes backups from 1999 to 2004, a range that is considerably broader than the 2002 time period sought by Plaintiff. In addition, 80-85% of the content is believed to be from systems other than the mail server. In addition, the 15-20% that comprises backups of the mail server would include many emails with no relevance to the present dispute and would contain the mailboxes of all CCBN employees (believe to be number 350), many with no involvement in the present dispute.

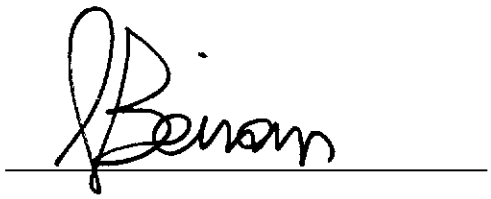
17. It is my understanding that Plaintiff has rejected these industry-standard processes for culling the collection and is seeking to have all of the contents of all of the tapes restored for production. To illustrate the volume of data involved, if all of the contents of all 478 tapes were

printed out as plain text at 5,000 characters per page, the estimated output would be 3 billion pages and would fill more than one million bankers boxes.

18. Similarly, if EED's pricing model were applied to all 478 tapes, and if all 16 terabytes of data were extracted from the tapes without culling at a rate of $1950 per gigabyte as the Plaintiffs request, the cost would be over $30 million. If EED's pricing model were applied only to the 15-20% that comprises the backups of the mail server, with no additional culling, the cost would still be approximately $5 million.

19. With such a large collection of tapes, even the simplest first step -- cataloging the tapes -- would be very expensive. Cataloging is necessary to provide a basis for determining which backups are actually recoverable, what they actually contain, whether any of the information on the tapes would be responsive to plaintiff's requests, and if so, how to cull the collection appropriately for cost-effective production. Therefore, cataloging is a prerequisite for any actual extraction, review, and production. At EED's rate of $250 per tape, cataloging would cost $119,500.

Signed under pains and penalties of perjury this 22nd day of September, 2005.



James Berriman